

# Advanced Genome Bioinformatics

Master in Bioinformatics UPF  
2016-2017

Eduardo Eyras  
Computational RNA Biology  
Pompeu Fabra University - ICREA  
Barcelona, Spain

## **Aims of the course**

- 1) Learn abstract and mathematical formulation of biological problems
- 2) Implement of a working program to study biological data
- 3) Critical evaluation of results and accuracy analysis of methods

## **Skills involved**

- 1) Probabilistic methods
- 2) Some algorithmics
- 3) Accuracy measures
- 4) Programming

## Course schedule and material

[http://comprna.upf.edu/courses/Master\\_AGB/](http://comprna.upf.edu/courses/Master_AGB/)

PowerPoint presentations

Exercises in class

## Evaluation

Assignment and presentation 60%

Written exam (theory) 40%

# Required

## Working Programs (perl/python)

1) Working program:

```
perl YourName_program.pl input.data options > output.data
```

```
perl -c YourName_program.pl
```

```
perl YourName_program.pl
```

} → should not give errors

2) Use: “**use strict**”

```
#!/usr/bin/perl -w
```

3) Program should provide usage instructions

```
perl YourName_program.pl
```

```
Usage: YourName_program.pl options...
```

# Required

4) Name your scripts consistently, e.g.:

**YourName\_AnalysisName.pl**

5) comment code (input, output, variables, algorithm steps, functions,...)

## **Presentation with:**

- 1) Methods
- 2) Results
- 3) Discussion

Send material to [eduardo.eyras@upf.edu](mailto:eduardo.eyras@upf.edu)

# OUTLINE OF THE COURSE

1. Probability and Classification
  - a. Bayesian classification
  - b. Accuracy analysis
  - c. Information and entropy
  - d. Description of known sequence motifs
  
2. Hidden Markov Models
  - a. HMM algorithms & examples
  - b. Profile HMMs
  
3. Iterative methods
  - a. Search for new motifs
  - b. Clustering

## Basic bibliography

1. Biological sequence analysis : probabilistic models of proteins and nucleic acids. R. Durbin, S. Eddy, A. Krogh, G. Mitchison. .
2. Problems and solutions in biological sequence analysis. M. Borodovsky and S. Ekisheva
3. Machine learning. Tom M. Mitchell
4. Methods for computational gene prediction. William H. Majoros
5. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. T. Hastie, R. Tibshirani & J. Friedman (<http://statweb.stanford.edu/~tibs/ElemStatLearn/>)
6. Elements of Information Theory. T. Cover & J. Thomas
7. Data Mining: Practical Machine Learning Tools and Techniques. M. Hall, I. Witten, E. Frank

## Complementary Bibliography:

1. Information Theory, Inference, and Learning Algorithms. D. McKay (<http://www.inference.phy.cam.ac.uk/itprnn/book.html>)
2. Combinatorial pattern matching algorithms in computational biology using Perl and R. G. Valiente
3. Genomic Perl : from bioinformatics basics to working code / Rex A. Dwyer      Dwyer, Rex A.
4. Introduction to algorithms. TH. Cormen, CE. Leiserson, RL. Rivest, C Stein
5. Machine learning approaches to bioinformatics. ZR Yang
6. An Introduction to bioinformatics algorithms. NC. Jones and PA. Pevzner
7. "Python Machine Learning", autor: Sebastian Raschka  
(<https://github.com/rasbt/python-machine-learning-book>)

## **Course schedule and material**

[http://comprna.upf.edu/courses/Master\\_AGB/](http://comprna.upf.edu/courses/Master_AGB/)